**Key findings**

**Go through content slides again**

* Beginners vs beginners-blockly
* Newbies vs beginners vs intermediate
* Present tables graphically – show how accuracy tends to decrease across challenges
* Horizontal – week # - lines (accuracy for different cases)
* Separate chart: tree sizes
* Try presenting in better report format ASAP – get results quite quickly
* **Predictive accuracy of no submissions** **was generally high for earlier challenges** (challenge-newbies-2018 and challenge-beginners-blockly-2018) but decreased for later challenges.
  + Valuable to predict no submission
* **Poor predictive accuracy of failed attempts due to low proportion of failed attempts** 
  + Task limited by availability of the data, so no way to solve the problem
  + In terms of useful data to predict failure better might be timeliness, but we still don’t have enough failure cases to train a classifier. We can speculate what’s useful information but not able to train.
* **Selection of problem slides over content slides:** Most of the slides used in the trees and selected by the feature selection algorithms were problem slides. Content slides played a minor role in most challenges, even in challenge-newbies-2018, where they were selected at least once every module.
  + **Flesh out the structure first and send across**
  + **TODO:** Look at specific content slides and speculate why they are so useful, potentially group, and send to Grok.
  + Give to educational team and discuss the value of the table. Provide a description.
  + Survey
  + Expert: These findings are useful because …
    - Incorporate parts of the transcript in thesis
  + Provide a list of content/problem slides selected and their scores
    - For this module, include **(select a few key problems and get feedback)**
      * the event that matters
      * the percentage contribution of the slides
      * provide decision tree
    - Get back a textual evaluation/interpretation of that information and some statements to the value of that information (qualitative description of the information)
    - In thesis, provide a digest of that.
    - Can provide a few questions (have a think – finalise asap)
      * Did you find this useful (rating)
      * Any surprises / new insights this gave you
      * Other
      * Include comments
  + For each problem, make a report card
    - Include the tree (just select one tree e.g. gain ratio) and how to interpret the tree
    - Make a note of the attributes selected and provide their scores
    - Predictors
    - Things you can use to evaluate the problem
  + Then talk about how useful this report is
* **Effect of feature selection methods in reducing size of tree:** Feature selection methods greatly reduced the number of leaves and the size of the tree, with negligible loss or even improvements to accuracy.

**challenge-newbies-2018:** All algorithms show consistently high predictive accuracy for predicting passed and no submission outcomes, as well as overall accuracy across all modules. All algorithms demonstrate a high predictive accuracy of no submissions (averaging at approx. 90%). All algorithms consistently fail to predict fail outcomes. The benefit of using feature selection is evident in the smaller and more interpretable tree. Both CFS and gain-ratio attribute-selection methods both outperform J48 wit feature selection, and demonstrate similarly high classifier performances.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **J48** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submission** | **Number of leaves** | **Size of tree** |
| **w1p1** | 86.68 | 85.14 | 25 | 91.61 | 15 | 28 |
| **w1p2** | 81.23 | 79.66 | 0 | 88.53 | 7 | 12 |
| **w2p1** | 85.47 | 85.38 | 0 | 85.83 | 5 | 8 |
| **w2p2** | 86.25 | 85.69 | 20 | 88.13 | 7 | 11 |
| **w3p1** | 86.05 | 83.19 | 0 | 93.57 | 3 | 4 |
| **w3p2** | 83.48 | 80.4 | 0 | 90.45 | 4 | 6 |
| **w4p1** | 86.76 | 80.63 | 0 | 93.69 | 6 | 9 |
| **w4p2** | 85.99 | 82.46 | 0 | 92.93 | 5 | 8 |
| **w5p1** | 85.16 | 78.00 | 0 | 95.79 | 3 | 4 |
| **w5p2** | 85.58 | 79.44 | 0 | 94.2 | 6 | 9 |
| **Average** | **85.265** | **82.00** | **4.5** | **91.473** | **6.1** | **9.9** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **BFS** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 86.14 | 83.55 | 0 | 94.85 | 6 | 10 |
| **w1p2** | 81.37 | 80.97 | 0 | 82.78 | 3 | 4 |
| **w2p1** | 85.47 | 85.24 | 0 | 86.06 | 4 | 6 |
| **w2p2** | 85.9 | 86.13 | 20 | 85.92 | 6 | 9 |
| **w3p1** | 86.05 | 83.19 | 0 | 93.57 | 3 | 4 |
| **w3p2** | 83.55 | 80.75 | 0 | 89.73 | 4 | 6 |
| **w4p1** | 86.83 | 80.73 | 0 | 93.69 | 6 | 9 |
| **w4p2** | 85.88 | 82.78 | 0 | 91.8 | 3 | 4 |
| **w5p1** | 85.16 | 78.00 | 0 | 95.79 | 3 | 4 |
| **w5p2** | 86.05 | 79.67 | 0 | 94.26 | 5 | 7 |
| **Average** | **85.24** | **82.10** | **2** | **90.845** | **4.3** | **6.3** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gain ratio** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submission** | **Number of leaves** | **Size of tree** |
| **w1p1** | 86.14 | 83.55 | 0 | 94.85 | 5 | 8 |
| **w1p2** | 81.34 | 79.6 | 0 | 88.43 | 4 | 6 |
| **w2p1** | 85.43 | 85.27 | 0 | 85.83 | 4 | 6 |
| **w2p2** | 86 | 86.35 | 20 | 85.6 | 5 | 7 |
| **w3p1** | 86.05 | 83.19 | 0 | 93.57 | 3 | 4 |
| **w3p2** | 83.55 | 80.75 | 0 | 89.73 | 4 | 6 |
| **w4p1** | 86.83 | 80.74 | 0 | 93.69 | 5 | 7 |
| **w4p2** | 85.88 | 82.78 | 0 | 91.8 | 3 | 4 |
| **w5p1** | 85.88 | 82.78 | 0 | 91.8 | 3 | 4 |
| **w5p2** | 86.05 | 79.67 | 0 | 94.26 | 5 | 7 |
| **Average** | **85.315** | **82.468** | **2** | **90.956** | **4.1** | **5.9** |

|  |  |  |  |
| --- | --- | --- | --- |
| **w1p1** | 0, 5, 10 | **Content**, problem, problem (missed first problem) | |
| **w1p2** | 5, 6, 10, 11 | Problem, problem, **content**, problem |  |
| **w2p1** | 3, 4, 6 | Problem, problem, **content** |  |
| **w2p2** | 3, 4, 5, 9 | Problem, problem, **content**, problem |  |
| **w3p1** | 5, 6, 7 | Problem, problem, **content** |  |
| **w3p2** | 4, 5, 13 | Problem, problem, problem |  |
| **w4p1** | 3, 5, 6 | **Content**, problem, problem |  |
| **w4p2** | 3, 7, 8 | **Content**, problem, problem |  |
| **w5p1** | 0, 5, 6, 9 | **Content**, problem, problem, content |  |
| **w5p2** | 4, 5, 8 | Problem, problem, problem |  |
|  |  |  |  |

Most modules had one content slide that was used selected as an attribute by either the CFS attribute-selection method or the gain-ratio attribute-selection method, except for two weeks (w3p2 and w5p2), which coincide with the mid-point and final point of the course.

**challenge-beginners-blockly-2018:** All algorithms show consistently high predictive accuracy for predicting passed and no submission outcomes, as well as overall accuracy across all modules, with the exception of w4p1 and w5p2. All algorithms demonstrate a high predictive accuracy of no submissions (averaging at approx. 87%). All algorithms consistently fail to predict fail outcomes. Like other challenges, the benefit of using CFS and gain-ratio is evident in the smaller and more interpretable tree (with small number of leaves and tree size). CFS outperforms gain-ratio and J48 without feature selection in predicting no submissions and passes, and has higher overall accuracy. (compare to baseline)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **J48** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 94.87 | 98.33 | 3.13 | 95.26 | 14 | 23 |
| **w1p2** | 92.62 | 98.42 | 12.5 | 89.92 | 8 | 13 |
| **w2p1** | 92.84 | 98.89 | 0 | 87.02 | 4 | 6 |
| **w2p2** | 87.79 | 98.18 | 0 | 79.18 | 8 | 12 |
| **w3p1** | 90.93 | 96.91 | 0 | 82.09 | 12 | 19 |
| **w3p2** | 89.64 | 95.5 | 0 | 92.01 | 8 | 13 |
| **w4p1** | 71.14 | 93.88 | 0 | 71.37 | 5 | 7 |
| **w4p2** | 91.42 | 95.62 | 0 | 90.59 | 6 | 9 |
| **w5p1** | 87.88 | 90.79 | 0 | 90.53 | 7 | 11 |
| **w5p2** | 89.18 | 98.09 | 10 | 91.73 | 4 | 6 |
| **Average** | **88.831** | **96.461** | **2.563** | **86.97** | **7.6** | **11.9** |
|  |  |  |  |  |  |  |
| **CFS** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 94.75 | 98.32 | 0 | 98.15 | 4 | 6 |
| **w1p2** | 92.51 | 98.42 | 12.5 | 89.92 | 4 | 6 |
| **w2p1** | 92.91 | 98.89 | 0 | 87.24 | 3 | 4 |
| **w2p2** | 88.05 | 98.18 | 0 | 79.86 | 3 | 4 |
| **w3p1** | 90.22 | 98.53 | 0 | 77.54 | 7 | 10 |
| **w3p2** | 90.24 | 95.69 | 0 | 93.4 | 5 | 7 |
| **w4p1** | 71.5 | 95.08 | 0 | 71.57 | 5 | 7 |
| **w4p2** | 91.42 | 95.62 | 0 | 90.59 | 3 | 4 |
| **w5p1** | 88.81 | 90.79 | 0 | 92.04 | 5 | 7 |
| **w5p2** | 88.65 | 98.09 | 5 | 91.34 | 3 | 4 |
| **Average** | **88.906** | **96.761** | **1.75** | **87.165** | **4.2** | **5.9** |
| **Gain ratio** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 94.66 | 97.92 | 0 | 95.31 | 3 | 4 |
| **w1p2** | 91.85 | 98.42 | 1.56 | 89.11 | 3 | 4 |
| **w2p1** | 92.98 | 92.91 | 0 | 87.47 | 3 | 4 |
| **w2p2** | 87.96 | 98.18 | 0 | 79.63 | 3 | 4 |
| **w3p1** | 90.22 | 98.53 | 0 | 77.54 | 7 | 10 |
| **w3p2** | 90.12 | 95.69 | 0 | 93.06 | 5 | 7 |
| **w4p1** | 71.14 | 94.69 | 0 | 70.99 | 5 | 7 |
| **w4p2** | 91.42 | 95.62 | 0 | 90.59 | 3 | 4 |
| **w5p1** | 88.81 | 90.79 | 0 | 92.04 | 5 | 7 |
| **w5p2** | 69.39 | 28.57 | 0 | 91.73 | 2 | 3 |
| **Average** | **86.855** | **89.132** | **0.156** | **86.747** | **3.9** | **5.4** |

|  |  |  |  |
| --- | --- | --- | --- |
| **w1p1** | 3, 7, 9, 13, 15 | Problem, problem, problem, problem, problem | |
| **w1p2** | 1, **3**, 7, 12 | Problem, **content**, problem, problem | |
| **w2p1** | 3, 5, 7, 11 | Problem, **content**, problem, problem | |
| **w2p2** | 2, 3, 6 | Problem, problem, problem |  |
| **w3p1** | 3, 9, 12, 14 | Problem, problem, problem, problem | |
| **w3p2** | 2, 7, 11 | Problem, problem, problem |  |
| **w4p1** | 2, 3, 4, 5, 13 | Problem, **content**, problem, **content**, problem | |
| **w4p2** | 3, 6, 10 | Problem, problem, problem |  |
| **w5p1** | 7, 12 | Problem, problem |  |
| **w5p2** | 4, 5 | Problem, problem |  |

Content slides are not selected much by the feature selection methods, appearing at least once on six different weeks across the challenge (w1p1, w1p2, w3p1, w3p2, w4p1 and w5p1).

**challenge-beginners-2018:** All algorithms show consistently high predictive accuracy for predicting passed and no submission outcomes, as well as overall accuracy across all modules, with the exception of w4p1 and w5p2. All algorithms demonstrate a relatively high predictive accuracy of no submissions (averaging at approx. 77%). All algorithms consistently fail to predict fail outcomes. Like other challenges, the benefit of using CFS and gain-ratio is evident in the smaller and more interpretable tree (with small number of leaves and tree size). All three algorithms have similar performances across all accuracy measures.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **J48** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 95.37 | 98.82 | 0.86 | 87.91 | 20 | 34 |
| **w1p2** | 95.37 | 98.75 | 13.48 | 85.12 | 18 | 31 |
| **w2p1** | 94.97 | 98.74 | 4.65 | 81.54 | 16 | 27 |
| **w2p2** | 93.16 | 98.6 | 8.97 | 76.22 | 14 | 24 |
| **w3p1** | 91.82 | 98.09 | 0 | 69.71 | 35 | 63 |
| **w3p2** | 92.86 | 98.3 | 5.21 | 81.4 | 5 | 7 |
| **w4p1** | 77.1 | 96.8 | 0 | 52.46 | 3 | 4 |
| **w4p2** | 90.22 | 97.98 | 7.89 | 77.47 | 12 | 20 |
| **w5p1** | 91.7 | 96.29 | 16 | 85.49 | 10 | 17 |
| **w5p2** | 85.14 | 97.21 | 1.83 | 78.85 | 6 | 9 |
| **Average** | **90.771** | **97.958** | **5.889** | **77.617** | **13.9** | **23.6** |
|  |  |  |  |  |  |  |
| **CFS** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 95.4 | 98.71 | 0.86 | 88.69 | 4 | 6 |
| **w1p2** | 95.18 | 98.79 | 6.74 | 84.4 | 5 | 8 |
| **w2p1** | 95.13 | 98.69 | 4.65 | 82.91 | 7 | 11 |
| **w2p2** | 93.04 | 98.18 | 6.41 | 77.53 | 5 | 7 |
| **w3p1** | 91.66 | 99.06 | 0 | 65.01 | 7 | 10 |
| **w3p2** | 92.48 | 97.59 | 5.21 | 82.26 | 5 | 7 |
| **w4p1** | 77.25 | 96.73 | 0 | 52.94 | 3 | 4 |
| **w4p2** | 90.31 | 97.51 | 10.53 | 79.19 | 5 | 7 |
| **w5p1** | 91.96 | 96.06 | 0 | 86.88 | 3 | 4 |
| **w5p2** | 85.14 | 97.15 | 0 | 79.05 | 3 | 4 |
| **Average** | **90.755** | **97.847** | **3.44** | **77.886** | 4.7 | 6.8 |
|  |  |  |  |  |  |  |
| **Gain ratio** | **Overall Accuracy** | **Accuracy Passed** | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 95.34 | 98.62 | 4.31 | 88.33 | 3 | 4 |
| **w1p2** | 94.86 | 98.02 | 10.11 | 85.84 | 7 | 10 |
| **w2p1** | 94.96 | 98.49 | 4.65 | 82.91 | 7 | 10 |
| **w2p2** | 93.09 | 98.18 | 14.7 | 77.35 | 5 | 7 |
| **w3p1** | 91.66 | 99.06 | 0 | 65.01 | 7 | 10 |
| **w3p2** | 92.4017 | 97.52 | 5.21 | 82.16 | 5 | 7 |
| **w4p1** | 77.25 | 96.73 | 0 | 52.94 | 3 | 4 |
| **w4p2** | 90.31 | 97.51 | 10.53 | 79.19 | 5 | 7 |
| **w5p1** | 91.93 | 96.07 | 4 | 86.73 | 5 | 7 |
| **w5p2** | 85.14 | 97.15 | 0 | 79.05 | 3 | 4 |
| **Average** | **90.69417** | **97.735** | **5.351** | **77.951** | **5** | **7** |

Unlike challenge-beginners-newbies-2018, content slides are not selected much at all by the feature selection methods, applied on challenge-beginners-blockly-2018. Content slides only appear in four modules modules: w1p2, w2p1, w2p2 and w4p1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **w1p1** | 16, 19 | Problem, problem |  |  |
| **w1p2** | 10, 14, 15 | **Content, content**, problem (first two problems not included) | |  |
| **w2p1** | 6, 8, 12, 13 | **Content**, problem, problem, content (first problem not included) | |  |
| **w2p2** | 7, 9, 11, 14 | Problem, **content**, problem, problem (first problem not included) | |  |
| **w3p1** | 3, 9, 12 | Problem, problem, problem |  |  |
| **w3p2** | 2, 7, 12 | Problem, problem, problem |  |  |
| **w4p1** | 2, 4, 8, 10 | Problem, problem, **content, content** |  |  |
| **w4p2** | 4, 8, 12 | Problem, problem, problem |  |  |
| **w5p1** | 7, 12, 13 | Problem, problem, content |  |  |
| **w5p2** | 4, 5 | Problem, problem |  |  |

**challenge-intermediate-2018:** All algorithms show consistently high predictive accuracy for predicting passed outcomes, as well as overall accuracy across all modules. All algorithms demonstrate a reduced predictive accuracy of no submissions (averaging at approx. 72%) compared to other challenges, which average >85%. There is also much greater variation in accuracy of no submissions, with cases such as w2p1 – w3p1 and w4p2 that decrease accuracy of no submission. All algorithms consistently fail to predict fail outcomes. Like other challenges, the benefit of using CFS and gain-ratio is evident in the smaller and more interpretable tree (with small number of leaves and tree size). All three methods (J48, CFS and gain-ratio) demonstrate comparable performance across all accuracy metrics. , and has higher overall accuracy. (compare to baseline)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **J48** | **Overall Accuracy** | **Accuracy Passed** |  | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 91.49 | 98 |  | 0 | 71.14 | 10 | 18 |
| **w1p2** | 93.88 | 97.73 |  | 2.67 | 81.56 | 10 | 17 |
| **w2p1** | 88.69 | 98.63 |  | 2.86 | 58.12 | 9 | 14 |
| **w2p2** | 88.39 | 97.64 |  | 0 | 65.06 | 14 | 26 |
| **w3p1** | 88.47 | 97.19 |  | 0 | 63.58 | 10 | 18 |
| **w3p2** | 91.11 | 97.47 |  | 0 | 77.23 | 7 | 12 |
| **w4p1** | 85.26 | 95.28 |  | 0 | 74.02 | 8 | 14 |
| **w4p2** | 79.55 | 96.44 |  | 0 | 63.72 | 3 | 4 |
| **w5p1** | 95.26 | 97.61 |  | 0 | 90.45 | 4 | 6 |
| **w5p2** | 86.93 | 95.74 |  | 0 | 79.63 | 3 | 4 |
| **Average** | **88.903** | **97.173** |  | **0.553** | **72.451** | **7.8** | **13.3** |
|  |  |  |  |  |  |  |  |
| **CFS** | **Overall Accuracy** | **Accuracy Passed** |  | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 91.6 | 97.97 |  | 0 | 71.76 | 3 | 4 |
| **w1p2** | 93.99 | 97.68 |  | 1.33 | 82.83 | 4 | 6 |
| **w2p1** | 87.97 | 97.39 |  | 0 | 59.71 | 5 | 7 |
| **w2p2** | 88.53 | 97.02 |  | 0 | 67.58 | 3 | 4 |
| **w3p1** | 88.59 | 96.96 |  | 0 | 64.91 | 3 | 4 |
| **w3p2** | 91.27 | 97.42 |  | 0 | 77.99 | 3 | 4 |
| **w4p1** | 85.26 | 95.11 |  | 0 | 74.33 | 3 | 4 |
| **w4p2** | 79.55 | 96.44 |  | 0 | 63.72 | 3 | 4 |
| **w5p1** | 95.55 | 97.88 |  | 0 | 90.79 | 4 | 6 |
| **w5p2** | 86.93 | 95.74 |  | 0 | 79.8 | 3 | 4 |
| **Average** | **88.924** | **96.961** |  | **0.133** | **73.342** | **3.4** | **4.7** |
|  |  |  |  |  |  |  |  |
| **Gain ratio** | **Overall Accuracy** | **Accuracy Passed** |  | **Accuracy Fail** | **Accuracy No Submision** | **Number of leaves** | **Size of tree** |
| **w1p1** | 91.62 | 97.97 |  | 0 | 71.85 | 3 | 4 |
| **w1p2** | 94.01 | 97.68 |  | 0 | 83.15 | 3 | 4 |
| **w2p1** | 87.97 | 97.39 |  | 0 | 59.71 | 5 | 7 |
| **w2p2** | 88.53 | 96.99 |  | 0 | 67.58 | 3 | 4 |
| **w3p1** | 88.59 | 96.96 |  | 0 | 64.91 | 3 | 4 |
| **w3p2** | 91.27 | 97.42 |  | 0 | 77.99 | 3 | 4 |
| **w4p1** | 85.26 | 95.11 |  | 0 | 74.33 | 3 | 4 |
| **w4p2** | 79.55 | 96.44 |  | 0 | 63.72 | 3 | 4 |
| **w5p1** | 95.4 | 97.14 |  | 0 | 92.13 | 3 | 4 |
| **w5p2** | 86.93 | 95.74 |  | 0 | 79.8 | 3 | 4 |
| **Average** | **88.913** | **96.884** |  | **0** | **73.517** | **3.2** | **4.3** |

Unlike challenge-beginners-newbies-2018, content slides are not selected much at all by the feature selection methods, applied on challenge-beginners-blockly-2018. Content slides only appear in three modules: w1p2, w2p1 and w4p1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **w1p1** | 1, 7, 8 | **Content, content**, problem | |  |
| **w1p2** | 11, 13 | Problem, **content** (first problem not included) | |  |
| **w2p1** | 3, 11 | Problem, problem |  |  |
| **w2p2** | 3, 15 | Problem, problem |  |  |
| **w3p1** | 1, 11 | **Content**, problem |  |  |
| **w3p2** | 5, 16 | **Content**, problem (first problem not included) | |  |
| **w4p1** | 0, 5 | **Content**, problem |  |  |
| **w4p2** | 5 | Problem |  |  |
| **w5p1** | 1, 10, 11 | **Content**, problem, **content** | |  |
| **w5p2** | 8 | Problem |  |  |

Swap intermediate and beginners-blockly descriptions